4К

УДК 004.428.4:519.651.2

А.А. Каленчук-Порханова, Л.П. Вакал
Институт кибернетики им. В.М. Глушкова НАН Украины, г. Киев
ioanna@public.icyb.kiev.ua

Аппарат аппроксимации в составе программного обеспечения суперкомпьютера с кластерной архитектурой

В статье описаны разработанные авторами алгоритмы и программные комплексы наилучшей равномерной аппроксимации с обоснованием их преимуществ. Для иллюстрации рассмотрены некоторые примеры их эффективного применения на практике, в том числе на суперкомпьютере с кластерной архитектурой.

Введение

На современном этапе развития уровень информационного обеспечения является одним из определяющих факторов развития экономики, науки, техники, и можно утверждать, что от количества и качества полученной информации существенно зависит эффективность жизнедеятельности общества в целом.

На практике информация, как правило, представляется в виде массивов числовых данных, которые являются дискретным представлением функциональных зависимостей, характеризующих исследуемые объекты и процессы разной природы. Работа с такими массивами связана с рядом серьёзных трудностей, возникающих при их использовании, например, в задачах математического моделирования; при восстановлении значений дискретно заданной функции на «неосвещенных» замерами участках, а также при хранении и скоростной передаче по каналам связи больших и сверхбольших по объему массивов. Для преодоления указанных трудностей применяется математическая обработка массивов числовых данных с использованием аппарата приближения (аппроксимации) функций с целью сжатия этих массивов путём замены дискретного представления функциональных зависимостей аналитическими выражениями (аппроксимантами) с небольшим числом параметров-коэффициентов. Степень сжатия характеризуется коэффициентом сжатия C, который определяется по формуле [1]:

$$C = b(f)/b(F), \tag{1}$$

где $b(f)$, $b(F)$ – число бит, необходимых для хранения функции f и аппроксиманта F.

Качественно новым способом такой замены является способ наилучшей равномерной (чебышевской) аппроксимации, который значительно эффективнее и универсальнее интерполяционного и среднеквадратического способов приближения. Главным преимуществом чебышевского способа является обеспечение точности приближения, полученной на некотором подмножестве точек промежутка аппроксимации, во всех точках этого промежутка. Кроме того, наилучшее равномерное приближение даёт лучшую точность аппроксимации по сравнению с наилучшим среднеквадратическим приближением аппроксимантами тех же классов.

В Институте кибернетики разработаны интеллектуализированные методы, алгоритмы и программные комплексы построения наилучших чебышевских аппроксимантов разных классов (полиномы, дробно-рациональные, экспоненциальные, логарифмические выражения и др.) [2-5]. Разработаны также алгоритмы и программы наилучшего кусочно-полиномиального приближения с разбиением массивов на группы элементов (сегментная аппроксимация) [6], а также чебышевского приближения функций многих переменных [7].

Аппарат наилучшей равномерной аппроксимации эффективно используется для сжатия больших массивов данных при решении актуальных проблем, связанных с расчётом характеристик сложных динамических систем, которые требуют высокой точности результатов, а также для решения задач математического моделирования и прогнозирования процессов разной природы, в частности, экологических процессов.

В настоящее время для суперкомпьютера с кластерной архитектурой СКИТ Института кибернетики разработаны библиотека программ наилучшей чебышевской аппроксимации и библиотека процедур вычисления математических функций с использованием аппарата наилучшей равномерной аппроксимации. В стадии разработки в составе программного обеспечения СКИТ находится пакет аппроксимации функций, который включает в себя программные комплексы всех способов аппроксимации. Создана также подсистема сжатия больших массивов числовых данных в составе Информационно-аналитической системы «Бюджетный комитет» с целью прогнозирования основных макроэкономических показателей бюджета Украины и обеспечения скоростной передачи по компьютерным сетям файлов размерностью в десятки мегабайт.

Постановка задачи, методы и алгоритмы её решения

Наилучшим равномерным приближением с весом $w(x)\neq 0$ для функции $f(x)$ на множестве точек E, где $E=[\alpha,\beta]$ или $E_N=\{x_1,x_2,\ldots,x_N\}\subset[\alpha,\beta]$, называется такой аппроксимант $F^*(x;A)$ из заданного класса функций {F(x;A)}, A = ($\alpha_0,\alpha_1,\ldots,\alpha_n$) и $n<N$, для которого выполняется условие

$$\max_{x\in E}\left|w(x)\left[f(x)-F^*(x;A)\right]\right|\equiv\rho=\min_{A}\max_{x\in E}\left|w(x)\left[f(x)-F(x;A)\right]\right|. \qquad (2)$$

Величина ρ называется величиной наилучшего равномерного взвешенного приближения. При $w(x)=1$ имеем наилучшее абсолютное, а при $w(x)=1/f(x)$ – наилучшее относительное приближение.

Наиболее известным и эффективным методом построения наилучшего равномерного приближения является метод последовательных чебышевских интерполяций (п.ч.и) Ремеза, который первоначально был разработан для случая аппроксимации полиномами $P_n(x)=\sum_{i=0}^{n}a_ix^i$, а затем распространён и на ряд других случаев [8]. Теоретической основой метода п.ч.и. является теорема Чебышева, в соответствии с которой полином наилучшего приближения характеризуется таким необходимым и достаточным условием чебышевского альтернанса: на множестве точек $E_N\subset[\alpha,\beta]$ существуют по крайней мере $n+2$ точки $\alpha\le x_0\le x_1\le\ldots\le x_{n+1}\le\beta$, в которых уклонение функ-

ции от аппроксиманта $\Delta(x) = w(x)[f(x) - P_n(x)]$ достигает своего модуль-максимума ρ с чередованием знака. Тогда $P_n(x)$ – полином наилучшего приближения, а $\{x_i,\}_{i=0}^{n+1}$ – чебышевский альтернанс.

Метод Ремеза является итерационным и заключается в построении последовательности $(n+2)$ – точечных наборов $S_j = \left\{x_0^{(j)}, x_1^{(j)}, \ldots, x_{n+1}^{(j)}\right\} \subset E_N$, $j = 1,2,\ldots$, сходящейся к чебышевскому альтернансу. При этом на каждом j-м шаге $n+1$ коэффициент текущего аппроксиманта $P_{n,j}(x)$ и погрешность приближения ρ_j определяются из системы линейных уравнений

$$w\left(x_i^{(j)}\right) \cdot \left[f\left(x_i^{(j)}\right) - P_{n,j}\left(x_i^{(j)}\right)\right] = (-1)^i \rho_j, \; i = \overline{0, n+1}. \quad (3)$$

Фактическая скорость сходимости п.ч.и. зависит, в основном, от различных способов замены наборов S_j, а именно – допустимого, полуоптимального и оптимального соответственно с такими скоростями сходимости: линейной, геометрической прогрессии и квадратической. В разработанных авторами алгоритмах, базирующихся на методе п.ч.и., реализуется предложенный в работе [9] полуоптимальный (на практике совпадающий с оптимальным) способ замены наборов точек S_j, квадратическая скорость сходимости которого обеспечивает нахождение наилучшего аппроксиманта, как правило, всего за 1 – 2 итерации, в то время как при других способах замены число итераций во много раз больше [8].

Ниже приводятся краткие описания разработанных в Институте кибернетики алгоритмов наилучшей чебышевской аппроксимации для функций как одной, так и многих переменных. Алгоритмы основаны на методе п.ч.и. Ремеза.

Описание алгоритмов аппроксимации

Алгоритмы п.ч.и. можно применять для аппроксимации как дискретно заданной, так и аналитически заданной функции. Во втором случае дополнительно вводится процедура вычисления значений функции в точках дискретизации.

Численная реализация предлагаемых алгоритмов построения наилучших приближений аппроксимантами различных типов имеет, кроме указанных выше, также дополнительные преимущества, связанные с оптимизацией этих алгоритмов по точности и быстродействию [3]. Для алгоритмов наилучшей чебышевской аппроксимации и соответствующих программных комплексов авторами получены оценки всех видов погрешностей, в частности, априорные и апостериорные мажорантные детерминированные оценки полной погрешности, причем неулучшаемые для некоторых классов функций [2]. Эти меры оптимизации позволили значительно повысить точность результатов вычислений (в некоторых случаях на порядок).

В свою очередь алгоритмы аппроксимации функций *одной переменной*, кроме оптимального варианта замены точек наборов S_j, имеют ряд возможностей и преимуществ по сравнению с алгоритмами п.ч.и. других авторов, а именно, обеспечивают построение полиномиального приближения с произвольным весом $w(x) \neq 0$; позволяют получить более точную оценку величины наилучшего приближения ρ за счет включения в вычислительные схемы алгоритмов расчёта полной погрешности приближения, а также позволяют находить либо аппроксимант заданной фиксированной степени (вход по степени), либо такой аппроксимант, который обеспечивает заданную точность приближения (вход по точности).

В алгоритме наилучшего чебышевского приближения полиномами вида $P_n(x)=\sum_{i=0}^{n} a_i x^i$ вычисление значений полинома в точках производится по схеме Горнера, а именно, $P_n(x)=((\ldots(a_n x + a_{n-1})x + \ldots)x + a_1)x + a_0$. В случаях, когда величина $2^{-\tau}\sum_{i=0}^{n}|a_i|$ погрешности округления коэффициентов a_i в системе с плавающей запятой с τ двоичными разрядами слишком велика, аппроксимант берётся в виде $P_n(x)=\sum_{i=0}^{n} c_i T_i(x)$, где $T_i(x)$ – многочлены Чебышева 1-го рода, а его значения в точках вычисляются по схеме Бахвалова [10]. Проведенный анализ показал, что для степеней аппроксимантов не выше 10 схемы Горнера и Бахвалова приблизительно равнозначны.

В случае *приближения дробно-рациональными выражениями* вида

$$R_{mk}(x)=\sum_{i=0}^{m} a_i x^i \Bigg/ \sum_{j=0}^{k} b_j x^j , \tag{4}$$

в отличие от аппроксимации полиномами, сходимость чебышевских интерполяций теоретически доказана только при условии близости начального приближения к наилучшему аппроксиманту. Поэтому в этом случае применяется подход, объединяющий преимущества алгоритма п.ч.и. Ремеза и алгоритма Вернера [11], а именно, высокую скорость сходимости первого и сходимость с произвольного начального приближения второго. Этот подход был реализован в комбинированном алгоритме [4], в котором для получения дробно-рационального аппроксиманта (4) сначала применяется метод Ремеза до нарушения его сходимости. Для обеспечения сходимости этого алгоритма в работу вступает алгоритм Вернера для получения нового начального приближения.

Далее описываются алгоритмы построения наилучших чебышевских приближений нелинейными выражениями (экспоненциальными, логарифмическими и др.), которые сводятся посредством использования вспомогательных функций к алгоритмам нахождения полиномиальных приближений с соответствующими весами.

Алгоритм чебышевского приближения экспоненциальными выражениями вида

$$F_n(x;A)= a_0\, exp\left(a_1 x + \ldots + a_n x^n\right),\ a_0 \neq 0 \tag{5}$$

позволяет находить наилучшее *относительное* приближение функции $f(x)$, поскольку выражениями вида (5) обычно аппроксимируют функции, значения которых не меняют знак. Наилучшее относительное приближение функции $f(x)$ аппроксимантом (5) сводится к наилучшему абсолютному приближению вспомогательной функции $ln|f(x)|$ полиномом вида

$$P_n(x;B)= \sum_{i=0}^{n} b_i x^i \tag{6}$$

с последующим пересчетом коэффициентов по формулам:

$$a_0 = \frac{2\cdot \exp(b_0)\cdot \exp(\tilde{\rho})\cdot sign\,(f(x))}{1+\exp\,(2\cdot\tilde{\rho})},\ a_i = b_i\ \left(i=\overline{1,n}\right),$$

где $\tilde{\rho}$ – величина наилучшего полиномиального приближения [12].

Замечание. Точки альтернанса при приближении выражением (5) и полиномом (6) совпадают, а величины ρ и $\tilde{\rho}$ соответственно наилучших экспоненциального и полиномиального приближений равны (с точностью до бесконечно малых высших порядков).

Алгоритм равномерного приближения логарифмическими выражениями вида

$$F_n(x;A) = ln\left(a_0 + a_1 x + \ldots + a_n x^n\right) \qquad (7)$$

позволяет находить наилучшее *абсолютное* приближение, так как выражениями вида (7) обычно аппроксимируют плавно изменяющиеся функции, которые могут менять знак. Наилучшее абсолютное приближение для функции $f(x)$ аппроксимантом (7) сводится к наилучшему относительному приближению для вспомогательной функции $exp(f(x))$ полиномом вида (6) с последующим пересчетом коэффициентов по формуле:

$$a_i = \frac{b_i}{\sqrt{1-\tilde{\rho}^2}} \quad \left(i = \overline{0,n}\right).$$

Для случая приближения логарифмическими выражениями предыдущее замечание относительно чебышевского альтернанса и равенства величин ρ и $\tilde{\rho}$ также справедливо.

Алгоритм чебышевской аппроксимации корнем целой степени из полинома

$$F_n(x;A) = \sqrt[l]{a_0 + a_1 x + \ldots + a_n x^n} \qquad (8)$$

позволяет находить наилучшее *относительное* приближение для функции $f(x)$. Как и в случаях приближений (5) и (7), приближение нелинейным выражением (8) сводится к наилучшей относительной аппроксимации полиномом вида (6) вспомогательной функции $[f(x)]^l$ и к вычислению коэффициентов аппроксиманта (8) по формулам:

$$a_i = b_i \left[\frac{2}{(1+\tilde{\rho})^{1/l} + (1-\tilde{\rho})^{1/l}}\right]^l \quad \left(i = \overline{0,n}\right).$$

При этом величина ρ наилучшего относительного приближения аппроксимантом вида (8) равна $\tilde{\rho}/l$ (с точностью до бесконечно малых высших порядков) [12].

Алгоритм сегментной аппроксимации полиномами предполагает разбиение всего промежутка аппроксимации $[\alpha,\beta]$ на r сегментов и приближение функции $f(x)$ отдельно на каждом i-м сегменте $\left(i = \overline{1,r}\right)$ полиномом $P_n^{(i)}(x)$ наилучшей равномерной аппроксимации с величиной наилучшего приближения ρ_i [6]. Величина сегментного приближения $\hat{\rho}$ определяется по формуле

$$\hat{\rho}(\tau) = \max_{1 \le i \le r} \rho_i . \qquad (9)$$

Обозначим T совокупность разбиений $\tau = \{\alpha \le t_0 < t_1 < \ldots < t_r \le \beta\}$ промежутка $[\alpha,\beta]$ на r сегментов $[t_{i-1}, t_i]$ $\left(i = \overline{1,r}\right)$. Разбиение $\tau^* = \{\alpha \le t_0^* < \ldots < t_r^* \le \beta\}$, для которого

$$\hat{\rho}\left(\tau^*\right) = \min_{\tau \in T} \hat{\rho}(\tau), \qquad (10)$$

называется *оптимальным*, а точки $t_0^*, t_1^*, \ldots, t_r^*$ – *оптимальными узлами.*

Разработанный алгоритм кусочно-полиномиальной сегментной аппроксимации позволяет находить приближение с величиной $\hat{\rho}$ и узлами разбиения, которые практически совпадают с величиной $\hat{\rho}\left(\tau^*\right)$ и оптимальными узлами (10). Алгоритм состоит из двух этапов. На первом – определяется минимальное число узлов, при котором для величины сегментного приближения $\hat{\rho}$ будет выполняться неравенство $\hat{\rho} \le \varepsilon$,

где ε – заданная погрешность аппроксимации. На втором этапе определяются оптимальные узлы и соответствующее кусочно-полиномиальное приближение. При этом используемая в алгоритме процедура определения узлов разбиения является более эффективной, чем обычная процедура половинного деления промежутка приближения на сегменты.

Следует подчеркнуть, что применение сегментной аппроксимации для случаев замены больших массивов данных позволяет значительно повысить точность приближения. В следующем разделе будет показано преимущество сегментной аппроксимации по сравнению с аппроксимацией без разбиения на сегменты при сжатии больших массивов данных.

Алгоритм равномерного приближения функций $f(X)$ k *переменных* $X=\{x_1,\ldots,x_k\}$ позволяет находить наилучшее приближение с весом $w=1$ обобщёнными многочленами

$$F_n(X;Z)=\sum_{j=1}^{n} z_j\varphi_j(X), \tag{11}$$

где $Z=(z_1,\ldots,z_n)$, по системе линейно независимых базисных функций $\varphi_1(X),\ldots,\varphi_n(X)$ на множестве точек $E_N=\{X^{(1)},\ldots,X^{(N)}\}$. Задача наилучшей равномерной аппроксимации многочленами (11) решается как частный случай задачи построения наилучшего в чебышевском смысле приближения к решению системы несовместных линейных уравнений

$$\Phi_i(Z)=\sum_{j=1}^{n} z_j\varphi_j\left(X^{(i)}\right)-f\left(X^{(i)}\right),\ (i=\overline{1,N}), \tag{12}$$

т.е. задачи определения такой системы значений параметров $Z=(z_1,\ldots,z_n)$, чтобы

$$\max_{1\le i\le N}|\Phi_i(Z)|\equiv L(Z)\to\min_Z. \tag{13}$$

Присоединяя к каждой функции $\Phi_i(Z)$ её симметрическую копию $\Phi_{N+i}(Z)=-\Phi_i(Z)$, задачу (12) – (13) можно представить в виде задачи линейного программирования с $n+1$ неизвестными и $2N$ ограничениями:

$$\Lambda=\min, \tag{14}$$

$$\xi_i=\Lambda-\Phi_i(Z)\ge 0\ \left(i=\overline{1,2N}\right). \tag{15}$$

Разработанный алгоритм наилучшей чебышевской аппроксимации функции k переменных обобщенными многочленами (11) заключается в сведении задачи приближения к задаче линейного программирования с ведущей двойственной максимум-задачей, которая решается модифицированным симплекс-методом (м.с.-м.) с учетом того, что на практике число уравнений значительно больше числа неизвестных и таблица «расширенного базиса» при м.с.-м. существенно меньше опорной таблицы при использовании прямого симплекс-метода. Данный алгоритм по сравнению с аналогичными позволяет находить наилучшие приближения вида (11) с большей точностью и, как правило, за значительно меньшее количество итераций [7]. Этого удалось достигнуть за счёт усовершенствования вычислительной схемы алгоритма посредством применения приёмов, которые позволяют сократить почти вдвое стандартную симплекс-таблицу и в процессе решения двойственной задачи м.с.-м. преобразовывать только модифицированную (сжатую) симплекс-таблицу, оставляя при этом неизменной опорную таблицу.

Примеры применения аппарата аппроксимации

Эффективность разработанного авторами аппарата аппроксимации проверена на протяжении многих лет при решении задач математической обработки массивов числовых данных в разных областях науки и техники [5].

В настоящее время для отечественного суперкомпьютера с кластерной архитектурой СКИТ создана библиотека процедур вычисления элементарных, гиперболических и специальных функций, для нахождения наборов коэффициентов наилучших чебышевских приближений которых применялся данный аппарат аппроксимации. Следует заметить, что полученная при этом точность аппроксимации была *не ниже* 10^{-19} при количестве коэффициентов, как правило, *не более 10*, что значительно лучше по сравнению с существующими аппроксимациями другими способами приближений.

Для иллюстрации результатов применения аппарата аппроксимации приводятся также примеры сжатия больших массивов данных, которые были получены на СКИТ в рамках создания Информационно-аналитической системы «Бюджетный комитет». При этом удалось значительно повысить коэффициенты сжатия (1) за счёт применения сегментной аппроксимации. Например, в результате сжатия массивов (файлов) размерами 3.5 Мб, 4 Мб и 7.7 Мб их размеры уменьшились соответственно до 24 Кб, 30 Кб и 20 Кб с коэффициентами сжатия 125, 140 и 400.

Ниже на рис. 1 для иллюстрации преимуществ сегментной аппроксимации приводится диаграмма соотношения точностей аппроксимации при сжатии одномерного массива-вектора, состоящего из 2000 элементов, с одинаковыми коэффициентами сжатия соответственно для случаев целого промежутка и разбиения его на сегменты.

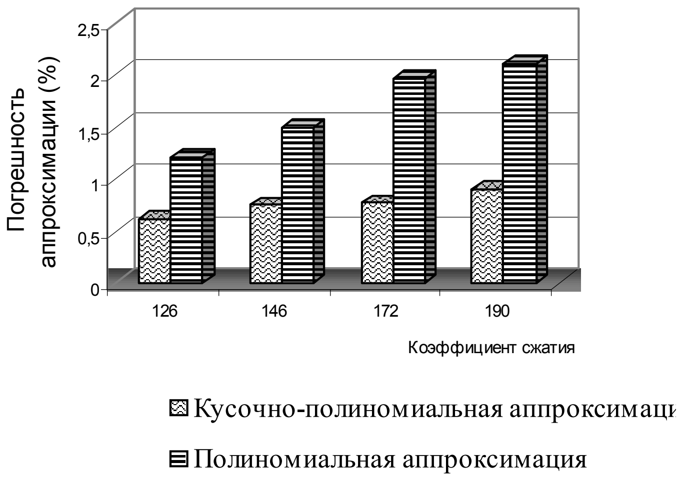


Рисунок 1 – Диаграмма сравнения погрешностей кусочно-полиномиального (сегментного) и полиномиального приближений при одинаковых коэффициентах сжатия

Выводы

Необходимость и важность проблемы аналитического представления потоков численной информации определяют особую актуальность разработки методов, алгоритмов и программных средств для его получения. Предлагаемый аппарат аппроксимации позволяет не только решить эту проблему, но и обладает рядом важных преимуществ, основанных на реализации способа равномерной аппроксимации, а именно,

обеспечивает получение наилучшего равномерного приближения как аналитически, так и дискретно заданных функций одной и многих переменных с использованием аппроксимантов различных классов с произвольным весом. Повышение эффективности применения данного аппарата достигается также за счёт оптимизации алгоритмов и программ по точности и быстродействию, что делает его предпочтительнее известных в литературе аналогичных работ. Использование на практике на протяжении многих лет программных реализаций разработанных алгоритмов аппроксимации на компьютерах различных типов, в том числе на суперкомпьютере с кластерной архитектурой, подтвердили их высокую эффективность.

Литература

1. Бердышев В.И., Петрак Л.В. Аппроксимация функций, сжатие численной информации, приложения. – Екатеринбург: УрО РАН, 1999. – 297 с.
2. Каленчук-Порханова А.А. Алгоритмы и анализ погрешности наилучшей чебышевской аппроксимации одной переменной // Труды Междунар. конф. теории приближения функций. – Москва – 1977. – С. 213-218.
3. Иванов В.В., Каленчук А.А. Об эффективности алгоритмов полиномиальных и дробно-рациональных чебышевских приближений // Конструктивная теория функций. – София, 1983. – С. 72-77.
4. Каленчук-Порханова А.А. Аппроксимация функций одной и многих переменных // Численные методы для многопроцессорного вычислительного комплекса ЕС. – М.: Изд-во ВВИА им. Н.Е. Жуковского, 1987. – С. 366-395.
5. Вакал Л.П., Каленчук-Порханова А.О. Аналітична обробка даних на основі чебишовської апроксимації // Математичні машини і системи. – 2006. – № 2. – С. 15-24.
6. Вакал Л.П. Рівномірне кусково-поліноміальне наближення // Комп’ютерні засоби, мережі і системи. – 2006. – № 5. – С. 53-59.
7. Каленчук-Порханова А.О., Вакал Л.П. Побудова найкращих рівномірних наближень функцій багатьох змінних // Комп’ютерні засоби, мережі і системи. – 2007. – № 6. – С. 141-148.
8. Ремез Е.Я. Основы численных методов чебышевского приближения. – К.: Наук. думка, 1969. – 623 с.
9. Каленчук-Порханова А.А. Об одном алгоритме полиномиальной чебышевской аппроксимации // Оптимизация вычислительных методов. – К.: Ин-т кибернетики АН УССР, 1974 . – С. 45-51.
10. Бахвалов Н.С., Жидков Н.П., Кобельков Г.М. Численные методы. – М.: Наука, 1987. – 600 с.
11. Werner H. Tschebysceff Approximation im Berich der Rationalen Funktionen bei Verliegen einer Guten Ausgansnäherung // Arch. Rat. Mech. Anal. – 1962. – Vol.10, № 3. – P. 205-219.
12. Попов Б.А. Равномерное приближение сплайнами. – К.: Наук. думка, 1989. – 272 с.

А.О. Каленчук-Порханова, Л.П. Вакал

Апарат апроксимації у складі програмного забезпечення суперкомп’ютера з кластерною архітектурою

У статті описано розроблені авторами алгоритми і програмні комплекси найкращої рівномірної апроксимації з обґрунтуванням їх переваг. Для ілюстрації розглянуто деякі приклади їх ефективного застосування на практиці, у тому числі на суперкомп’ютері з кластерною архітектурою.

A.A. Kalenchuk-Porkhanova, L.P. Vakal

Approximation Apparatus Composed of Software for Supercomputer with Cluster Architecture

In the paper algorithms and bundled software developed by authors for best uniform approximation are presented and ground of their advantages are given. Some examples of their practical application including their usage for supercomputer with cluster architecture are given.

Статья поступила в редакцию 30.09.2008.